

# The Effect of Duration and Glottalization on the Perception of Rhythm

NIAMH KELLY, MEGAN CROWHURST, CRYSTAL COBB

*The University of Texas at Austin*

## 1 Introduction

Humans are finely attuned to rhythm in many physical domains, from the purely anatomical to the perception of acoustic fluctuations in the speech signal. In fact, humans tend to hear rhythmic patterns even when sequences of sounds are physically indistinct, such as the clicking of a metronome or dripping water (e.g. Bolton 1894, Iversen et al. 2008, Fletcher 2010, Crowhurst and Teodocio 2014). Human languages exhibit rhythmic patterns expressed in the alternation of stressed and unstressed syllables (e.g. Bolton 1894; Halle 1973; Prince 1983; Hayes 1980, 1995, among many others). Researchers have therefore been keenly interested in understanding the perceptual factors that might underlie rhythmic distinctions in the speech domain. One significant body of research has investigated the physical correlates of stressed and unstressed syllables in various languages, and their consequences for the perception of stressed syllables. An early theme in the psychoacoustic research of the late 19<sup>th</sup> century, now seeing renewed interest, pertains to the perception of natural groupings of sounds. In this paper, we report the outcome of a study that is aligned with research in the second category. We studied the influence of varying vowel duration and glottalization on listeners' preferred syllable pairings in multisyllabic alternating sequences.

### 1.1 Rhythmic Grouping and the Iambic-Trochaic Law

Early research on the psychology of grouping produced the two generalizations in (1), now known to phonologists as the Iambic-Trochaic Law, or ITL (Bolton 1894; Woodrow 1909; Hayes 1995).

(1) *The Iambic-Trochaic Law* (adapted from Hayes 1995)

- a. *The intensity principle*: Elements contrasting in intensity naturally form groupings with initial prominence (trochees).
- b. *The duration principle*: Elements contrasting in duration naturally form groupings with final prominence (iambs).

Modern experimental studies using both nonspeech and speech-like stimuli have confirmed the intensity principle. Successful experiments using nonspeech stimuli have tested speakers of Japanese (Kusumoto and Moreton 1997; Iversen et al. 2008); English (Rice 1992; Kusumoto and Moreton 1997; Hay and Diehl 2007; Iversen et al. 2008); and French (Hay and

Diehl 2007). The key studies for speech have tested speakers of English (Hay and Diehl 2007; Crowhurst 2013; Crowhurst and Teodocio 2014), French (Hay and Diehl 2007; Bhatara et al. 2013), German (Bhatara et al. 2013), Spanish (Crowhurst 2013), and Zapotec (Crowhurst and Teodocio 2014). While the research of greatest relevance for the work reported here has used a forced choice subjective grouping methodology, Morgan et al (2013), using a serial recall study, found that English speakers were better able to remember a list of six loud and soft syllables organized in trochaic pairs than when they were presented as iambs.

The findings for duration have been more mixed and suggest that the perception of rhythm may be affected by one's native language. Some recent studies have reported a short-long grouping preference for multisyllabic sequences alternating in duration among speakers of English, French, German, Spanish (see references for these languages cited in the last paragraph), and Italian (Bion et al. 2011). Trainor and Adams (2000) report results suggesting that American English speaking adults and eight month olds perceived increased vowel duration as marking ends of short-short-long groupings. Hay and Diehl (2007) also found a short-long grouping preference with English- and French speakers who were tested with alternating non-speech sequences. In contrast, two studies have found grouping biases that conflict with the duration principle in (1b). Japanese speaking participants tasked with grouping duration-varying tonal sequences in Iversen et al. (2008) fell into three groups: one displaying a short-long preference; one with a long-short bias; and the third having no strong preference.<sup>1</sup> Among the studies using speech-like stimuli, Crowhurst and Teodocio (2014) found a clear preference for long-short groupings among the Zapotec-speaking participants in their study, when duration was the only parameter manipulated.

While the studies described above have been limited in testing the influence of varying intensity and duration singly, at least three modern studies have tested the joint influence of intensity and duration on grouping. Streeter (1978) found that cues to duration were more important than intensity in signaling prosodic boundaries at the phrasal level. Two more recent subjective grouping studies have found the opposite: when speakers of American English and Spanish (Crowhurst 2013) and speakers of American English and Zapotec (Crowhurst and Teodocio 2014) were tasked with grouping rhythmic multisyllabic sequences in which both intensity and duration were manipulated, the results suggested that intensity was a stronger predictor of listeners' grouping tendencies than was duration. These studies had different goals, however. The focus of Streeter (1978) was in relative salience of cues signaling phrase boundaries. On the other hand, the materials used in Crowhurst (2013) and Crowhurst and Teodocio (2014) (as well as subjective grouping studies described in the previous paragraphs) were meant to simulate binary rhythms that are more characteristic of iterative stress systems.

The conflicting effects found for duration as a predictor of listeners' grouping behavior, described above, suggest that the effect of a particular acoustic cue on the perception of natural

---

1 Similar studies, Kusumoto and Moreton (1997) and Iversen et al. (2008), found no reliable preference for either short-long or long-short groupings among their Japanese-speaking participants.

groupings may be influenced by the listener's native language. At present, however, the reasons for these differences are not well understood. The number of experimental investigations into grouping preferences is limited; in particular, those using speech-like stimuli are small in number. For this reason, we have no detailed typology of speech-related cues that can affect grouping behavior because (to our knowledge) no ITL-style studies with speech have tested the influence of cues other than intensity and duration on the perception of natural groupings. The current investigation contributes to this literature by testing the influence of vowel glottalization on subjective grouping preferences with speakers of American English. The influence of vowel duration was also tested, both singly, and in combination with glottalization. Section 1.2 presents the rationale for the choice of these features. The experiment and its findings are presented in section 2, followed by a discussion of these findings and their implications in section 3.

## **1.2 Duration and Glottalization in English**

In English, duration has been shown to be the most useful cue for the perception of stress (Fry 1955, Gay 1978). Likewise in production, duration is the most reliable cue to stress (Fry 1958, Lehiste 1970, Beckman and Edwards 1994, Campbell and Beckman 1997). Increased duration also has a demarcative function in signaling the end of constituents at all levels of the prosodic hierarchy in English and in many other languages (Gussenhoven and Rietveld 1992, Byrd et al. 2006). This phenomenon of final lengthening may well be associated with the psychoacoustic generalization in (1b).

In American and UK varieties of English, glottalization is also a demarcative feature that can signal endings of at least higher order prosodic constituents (Dilley et al. 1996, Redi and Shattuck-Hufnagel 2001). Glottalization (or laryngealization) can refer to the presence of a glottal stop or to creaky voicing. Creaky voice (also termed vocal fry) occurs when the vocal folds are tensed, or “tightly adducted but open enough along a portion of their length to allow for voicing” (Gordon and Ladefoged 2001:386). The relation between the glottal stop and creaky voice is that the presence of a glottal stop often induces creaky voice on neighboring sounds (Gordon and Ladefoged 2001). This occurs especially when the glottal closure is not complete, and instead the voicing pulses of the vocal folds are interrupted, resulting in creaky voice. Expanding on a previous categorization of glottalization by Huber (1988), Redi and Shattuck-Hufnagel (2001) describe four types, characterized according to their influence on the sound wave in speech: aperiodicity, creak, diplophonia, and glottal squeak. In an examination of glottalization in speakers of American English, Redi and Shattuck-Hufnagel (2001) found considerable variation both in terms of how much glottalization speakers used and in how it was realized. This study also found that in English, glottalization is more likely to occur at “prosodically significant locations such as phrase boundaries, utterance boundaries and pitch accents” (p. 408). American English speakers have been shown to produce more glottalization on word-initial vowels and on vowels with pitch accents (Pierrehumbert 1995), and at prosodic boundaries (Dilley et al. 1996). Glottalization can also be segmentally conditioned, being induced by surrounding glottal stops or final voiceless stops (Gordon and Ladefoged 2001).

As the experimental literature suggests that glottalization and prosody are interrelated, and in particular, that glottalization tends to signal boundaries, it is necessary to conduct further perception experiments to determine the extent to which listeners associate glottalization with finality. The current study contributes to this program of research by investigating whether the presence, and amount, of glottalization affects English speakers' grouping of syllables.

Given that both duration and glottalization have demarcative functions, and given English speakers tend to perceive longer vowels as group-final in alternating sequences, we asked whether the variations in vowel glottalization would also shape listeners' grouping preferences in a similar way. Our second question was whether one of these demarcative features would emerge as a more robust predictor of grouping preferences, when the two were co-varied in a way that forced listeners to intuitively choose between them.

## 2 The Experiment

In Experiment 1, native English speakers were exposed to speech streams in which the syllables *ba* and *ga* were alternated. Five types of sequence were included in the study. In two of these, a single feature, either vowel duration or vowel glottalization, was systematically varied. In the duration-varying set, length disparities were created by increasing the duration of the vowel in one alternating syllable (*ba* or *ga*) relative to that of the other at fixed ratios. In the glottalization-varying set, the vowel of one alternating syllable (*ga*) had a final period of creakiness. In the remaining sequences, glottalization and duration were both varied in one of two ways. In an in-phase set, the same syllables were marked by increased duration and glottalization: a short, modal syllable *ba* was alternated with a longer, creakier *ga*. In an out-of-phase set, a short, creaky *ga* was alternated with a longer, modal *ba*. Finally, the stimulus set included a “no difference” control sequence. Study participants were tasked with indicating whether they thought the sequences they heard consisted of recurrent *baga* or *gaba* syllable pairings.

### 2.1 Method

#### 2.1.1 Stimuli

Recordings were made of a male American English speaker producing sequences of *#BA-ga-BA-ga...* and *#GA-ba-GA-ba....* (in which capitalization indicates syllables that were emphasized naturally by the speaker). The recording took place in a sound-treated room at the University of Texas at Austin using a Shure close-talking microphone connected to a MOTU (Mark of the Unicorn) solid-state digital recorder. One clear, modal token of *ba* and one of *ga* were selected from the weaker position in the recorded sequences. The selected syllables chosen to be as closely matched as possible for pitch (both were between 97 and 101 Hz), intensity, and vowel quality. The vowels in the selected *ba* and *ga* syllables measured 217 and 250 ms, respectively.

The selected *ba* and *ga* were manipulated using standard functions in Praat (Boersma and Weenink, 2011) to produce variations in vowel duration and creakiness. Three versions of each syllable were produced, the duration of whose vowels were set to 150 ms, 200 ms, and 300 ms. Variations in duration were produced by copying and inserting, or removing full voicing cycles from the vowel in the original syllable, cutting at zero-crossing lines. Once length adjustments had been made, average intensity for all syllables was set to 65 dB by changing gain.

Glottalized versions of *ga* were created by synthesizing a creaky period at the end of the vowel.<sup>2</sup> This was done as described in Frazier (2009), by reducing every second pitch point in *ga* to 35 Hz. At level 1, the final 20% of the vowel was creaky, and at level 2, the final 50% was creaky. These proportions were constant, regardless of the absolute duration of the vowel, which also varied in the in-phase and out-of-phase conditions. (The absolute duration of the creaky period of any vowel can be calculated from the information provided in Table 1.) Figures 1 and 2 show the spectrograms for the creaky versions of *ga* whose vowels measured 300 ms.

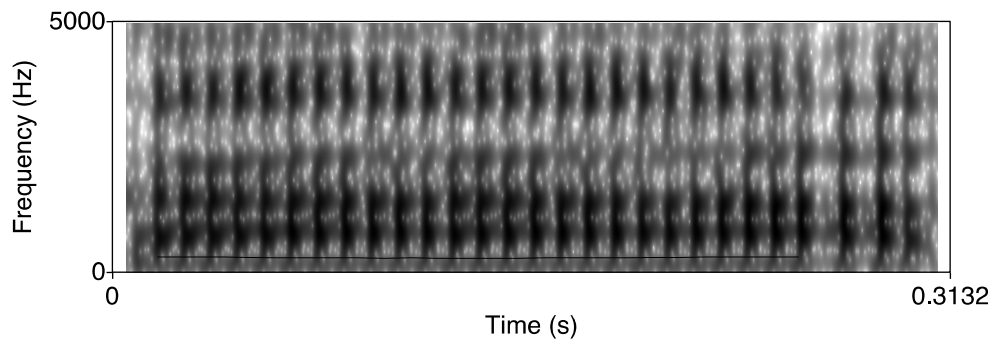


Figure 1: Long GA, 20% final creak (level 1)

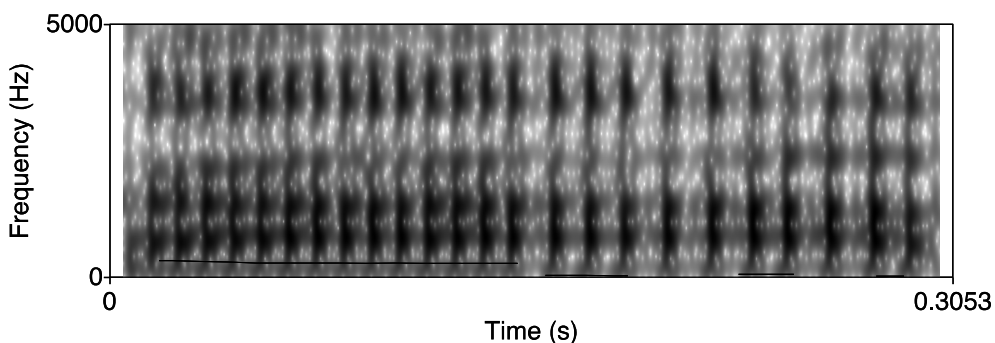


Figure 2: Long GA, 20% final creak (level 2)

---

2 While Redi and Shattuck-Hufnagel (2001) found aperiodicity to be the most common type of glottalization among the speakers examined, our synthesized creak was periodic. However, a drop in pitch has been found to be a reliable cue to the perception of glottalization (Dilley et al. 1996). Figures 1 and 2 show examples of the long syllables with final creak and half creak. Gerfen and Baker (2005) also found that for speakers of Coatzapan Mixtec, a drop in  $f_0$  was a sufficient cue for vowels to be perceived as laryngealized.

The syllables were then arranged into sequences alternating for duration and/or creakiness. Sequences of 10–11 seconds in duration were created by alternating a version of *ba* and of *ga*, with a 100 ms period of silence between syllables to simulate stop onsets. Sequences consisted of whole syllable pairs, either *baga* or *gaba*, to counterbalance string-initial and string-final syllables. The first 5 seconds of each sequence was combined with white noise that faded out from an amplitude of 63 dB to 0 dB, while the syllables faded in from 0 dB to 65 dB. The end of each sequence was followed by 500 ms of white noise at 65 dB to mask the sequence-final syllable. The design for the study, including the values assigned to the syllables *ba* and *ga* in the sequences used, is shown in Table 1. The final column in Table 1 indicates the actual grouping represented by a *baga* response in each condition.

	Manipulation levels		<i>baga</i> response =
	<i>ba</i>	<i>ga</i>	
Control (all modal)	D 200	D 200	( <i>ba ga</i> )
Duration varying (all modal)	D 150	D 300	( <i>ba gaa</i> )
	D 300	D 150	( <i>baa ga</i> )
Creak varying (modal + creaky)	D 200	Cr 1	( <i>ba ga?</i> )
	D 200	Cr 2	( <i>ba ga??</i> )
In-phase	D 150 x Cr 0	D 300 x Cr 1	( <i>ba gaa?</i> )
	D 150 x Cr 0	D 300 x Cr 2	( <i>ba gaa??</i> )
Out-of-phase	D 300 x Cr 0	D 150 x Cr 1	( <i>baa ga?</i> )
	D 300 x Cr 0	D 150 x Cr 2	( <i>baa ga??</i> )

Table 1: Experimental design. Parameters assigned to alternating syllables.  
(D = duration in ms, Cr = manipulation level for glottalization.)

### 2.1.2 Participants

The study participants were 24 male and female native speakers of English, aged 18-25, who had not had significant exposure to another language. These were recruited through a University-wide public events newsletter distributed by email, and they received \$10 for their time.

### 2.1.3 Task and Testing Procedure

Participants were tested in groups of 1 to 6 in a sound-treated room in a phonetics laboratory at the University of Texas at Austin. The experiment was controlled by SuperLab 4.0 software running on a MacBook Pro computer. Listeners heard sequences in free field, over a portable Bose SoundLink speaker connected by cable to the computer. Once the experimenter had conducted informed consent procedures and provided initial instructions, a four-trial test run was conducted to ensure that participants understood the task. The sequences were presented in 8 blocks of 12 trials, whose order was randomized by the software each time a block was run. The experimenter was present to manage the software for the duration of the experiment. Participants

were given response sheets containing numbered lines printed with arbitrary sequences of alternating *ba* and *ga* syllables, and were asked to circle or bracket any two adjacent syllables that matched the pairing they heard, either “baga” or “gaba” for each sequence. In this way we were able to obtain a uniform, indirect measure of listeners’ perceptions regardless of the type of sequence being evaluated and without the need for complicated explanations. They were also instructed to provide a confidence rating by circling “yes” or “no” to the right. A representative response sheet is shown in Appendix A. Including the initial instructions, the experiment took approximately 35 minutes to complete.

#### 2.1.4 Hypotheses

Given that both increased duration and vowel glottalization are associated with demarcation, we hypothesized that when either duration or glottalization was varied singly, the response data would reflect a bias favoring groupings in which the demarcative feature came last (that is, short-long and modal-creaky pairings). Given that the dependent variable measured *baga* responses, and that duration and glottalization were each varied on a 3-point scale, our specific hypotheses were those in (2a) and (2b).

- (2) a. *Baga* responses will decrease in proportion to increases in the duration of *ba* relative to that of *ga*.
- b. *Baga* responses will increase in proportion to the duration of the glottal segment in *ga*.

Hypothesis (2a) has been confirmed in prior research described in section 1, while (2b) has not been tested in earlier studies. We expected (2a) and (2b) to be confirmed in the in-phase condition as well, since here a *baga* response indicated both a modal-creaky and a short-long grouping. Since duration and glottalization cues were mutually reinforcing in the in-phase condition, we expected that the magnitude of the effect on the dependent variable in this condition would be greater overall than in either the duration-only or the glottalization-only condition. Therefore, we had the additional specific predictions in (3a) and (3b).

#### (3) In-phase condition

- a. The strength of the bias favoring *baga* groupings will be greater than in the duration-only condition.
- b. The strength of the bias favoring *baga* groupings will be greater than in the glottalization-only condition.

In the out-of-phase condition, in which a *baga* response simultaneously represented a long-short and a modal-creaky grouping, the hypotheses in (2a) and (2b) were in competition. We expected that in this condition, one of these hypotheses and not the other would be confirmed, but we had no reason to predict which cue, if any, would dominate.

## 2.2 Results

### 2.2.1 The Effect of Varying Duration and Glottalization Singly

As a first step in the analysis, trends in the response data for the control sequences and duration-only sequences (henceforth Set A) were examined.<sup>3</sup> The primary trend in the Set A data is represented in Figure 3 as the proportion of *baga* responses for points on the duration scale. In the control condition where *ba* and *ga* had the same duration, the proportion of *baga* responses was .41. At manipulation level -1 where *ba* was shorter than *ga*, the proportion of *baga* responses (representing a *bagaa* grouping) was higher, at .56; and at manipulation level 1 where *ba* was the longer syllable, the proportion of *baga* responses (a *baaga* grouping) was lowest, .37. Both outcomes indicate a bias favoring short-long groupings and together, they confirm the prediction in (2a).

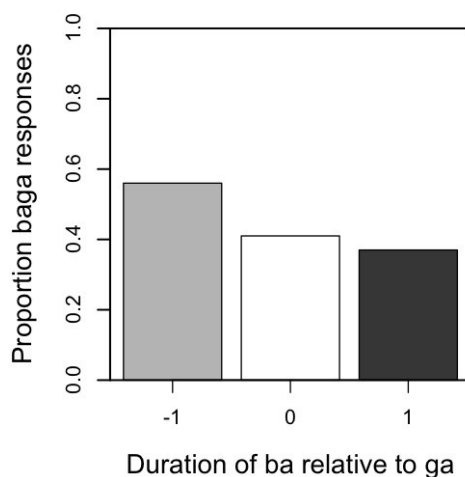


Figure 3: Proportion of *baga* responses in the control (white) and duration conditions (Set A).

To test the statistical significance of the trend associated with DURATION, a mixed-effects logistic regression model was fitted to the Set A data using the *lmer* function in the matrix library of the statistical software package R (Urbanek et al. 2012). The binary dependent variable measured the probability of choosing a *baga* response, which was arbitrarily coded as 1 in the analysis. SUBJECT was treated as a random variable. Consistent with the design in Table 1, DURATION was coded as a 3-point scale (-1, 0, 1), on which 0 represented the "no difference" control condition. Two design variables, BLOCK with 6 levels (one for each experimental block) and ORDER with 2 levels (depending on the string-initial syllable) were treated as factors in the analysis. The process of model selection followed the strategy of forward addition, which begins with a model that includes only the intercept. The model that provided the best fit for the Set A

---

3 The response data gathered in the experiment are given in Tables 6, 7, and 8 in Appendix.



data, whose output appears in Table 2, included terms for the intercept and for the predictor DURATION. This model provided a significantly better fit for the data than the intercept-only model, and was the most parsimonious model that optimized goodness of fit. Table 2 indicates that the effect of DURATION was highly significant. The odds ratio (the exponent of the estimated coefficient) provides a measure of effect size. The odds ratio associated with DURATION in Table 2 indicates that *baga* response was only .73 times as likely (or 27% less likely) per unit increase in the duration of *ba* relative to that of *ga*.

	Coef ( $\beta$ )	Odds	SE	z	P(> z )
INTERCEPT	-0.233	0.79	0.07	-3.035	= 0.002 **
DURATION	-0.315	0.73	0.07	-4.491	< 0.00001 ***

Table 2: Output of the best-fitting statistical model for the response data in Set A.

To explore the effect of varying glottalization singly, data Set B was prepared by combining response data for the control and glottalization-only sequences (see Table 7 in the Appendix). All vowels had the same duration (200 ms) in this condition. On the three point glottalization scale, level 0 represented a fully modal sequence, 1 represented sequences in which *ga* ended in a short period of creakiness (*gaʔ*), and 2 represented a sequence in which the second half of the vowel in *ga* was creaky, (*gaʔʔ*) (see Figures 1 and 2). Figure 4 reveals that the proportion of *baga* responses increased with increases in glottalization, as predicted by the hypothesis in (2b). However, while the presence of a creaky period increases *baga* responses, the difference between levels 1 and 2 was very small (.58 vs. .61). Pearson's chi-square test showed that this difference was not significant ( $X^2 = 0.443$ ,  $df = 1$ ,  $p = .51$ ).

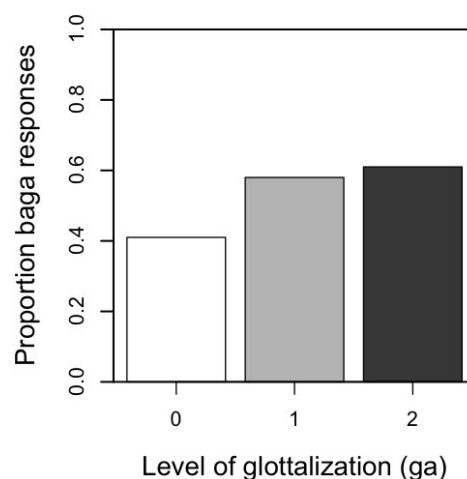


Figure 4: Proportion of *baga* responses in the control (white) and glottalization conditions (Set B).

As before, a mixed-effects logistic regression model was fitted to the Set B response data to test the statistical significance of the trend associated with glottalization. The statistical

procedures used were as described for Set A with the exception that the predictor variable was now CREAK, treated as a scale with 3 levels (0, 1, 2). The output of the best-fitting model is shown in Table 3.

	Coef ( $\beta$ )	Odds	SE	z	P(> z )
INTERCEPT	-0.317	.73	0.184	-1.724	= 0.0848
scale (CREAK)	0.469	1.60	0.090	5.244	< 0.00001***

Table 3: Output of the best-fitting statistical model for the response data in Set B.

This model, which included terms for the intercept and for CREAK, provided a significantly better fit for the data than the intercept-only model. No other variable contributed significantly to goodness of fit. The odds ratio associated with CREAK in Table 3 indicates that a *baga* response was 1.6 times more likely per unit of increase on the glottalization scale, an effect that was highly significant.

### 2.2.2 The Effect of Varying Duration and Glottalization Together

In the remaining two conditions, duration and glottalization were co-varied in the same sequences. In the in-phase sequences, a short, modal *ba* alternated with a longer, creakier *ga* (i.e. *gaa?* or *gaa??*). In the out-of-phase sequences, a long, modal *ba* (or *baa*) alternated with a short, creakier *ga* (i.e. *ga?* or *ga??*). The response data for these conditions occupy Table 8 in Appendix B.

We expected that in the in-phase condition, the proportion of *baga* responses would be higher than when either duration or glottalization was varied singly (see the hypotheses in 3a and 3b). However in the in-phase condition, a *baga* response represented both a short-long and a modal-creaky grouping. Should our prediction be confirmed, we were therefore also interested to know the relative contributions of duration and glottalization to the observed outcome. To examine these issues, data Set C was assembled from the response data for the duration-only subcondition in which *ga* was the longer syllable (manipulation level 1 in Figure 3), the glottalization-only condition, and the in-phase condition. The Set C results are charted in Figure 5. Labels beneath columns in the graph indicate properties of the groupings represented by a *baga* response in each condition and subcondition (“7” represents glottalization in Figures 5-7). In the comparison between *bagaa* (the duration-only sequence, white column) vs. *baga?* and *baga??* (the glottalization-only sequences, light grey), the chart reveals that a glottalization disparity produced more *baga* groupings than did a duration disparity. When *ga* was long, the proportion of *baga* responses was indeed higher when it was also creaky (dark grey columns) than when it was fully modal (the white column). And when *ga* was creaky, adding length (dark grey) increased *baga* responses in comparison to the glottalization-only condition (light grey), in which *ga* was short. However, consistent with the finding for the glottalization-only condition, the difference between the two in-phase subconditions was very small (proportion *baga* = .66, vs. .68). Again, Pearson’s chi-square test revealed that this difference did not reach significance

( $X^2 = 0.225$ ,  $df = 1$ ,  $p = .635$ ). Overall, Figure 5 reveals that our hypotheses in (3a) and (3b) were confirmed by the response data.

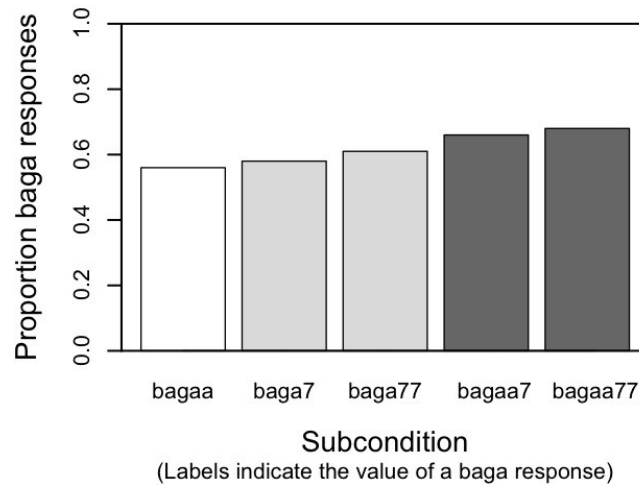


Figure 5: Proportion of *baga* responses in Set C. (White: duration-varying with long *ga*; Light grey: glottalization-only; Dark grey: in-phase)

The significance of any trends associated with duration and glottalization in the Set C data was tested using the procedures described for Sets A and B. As before, SUBJECT was treated as a random variable and the design variables ORDER and BLOCK were coded as factors. The predictor variables were CREAK coded as a 3 level scale (0-2), and GA.DURATION with two levels (0, 1), determined by the length of the syllable *ga*. The output of the best-fitting model is shown in Table 4. This model, which included terms for the intercept and both predictor variables, provided a significantly better fit for the data than models without either GA.DURATION or CREAK. Neither design variable contributed significantly to goodness of fit.

	Coef ( $\beta$ )	Odds	SE	z	P(> z )
INTERCEPT	0.379	1.46	0.183	2.072	= 0.0382*
GA.DURATION	0.298	1.35	0.124	2.400	= 0.0164*
scale (CREAK)	0.120	1.13	0.061	3.275	= 0.0011**

Table 4: Output of the best-fitting statistical model for the response data in Set C.

The statistical analysis revealed that the fixed effect of GA.DURATION was significant and that the effect of CREAK was highly significant. In other words, glottalization and duration did not contribute in equal proportion to increases in *baga* responses, a finding that is evident on close inspection of Figure 5 (in the comparison between the conditions represented by the white and light grey columns). We interpret this to mean that of the two features, glottalization was the stronger predictor of listeners' grouping biases. This interpretation is supported by the odds ratios associated with the terms GA.DURATION and CREAK.

Turning to the out-of-phase sequences, glottalization and increased duration were marked on different syllables in this condition, and for this reason, our specific hypotheses in (2a) and (2b) were in competition. The question of interest was whether one of these cues would prove to be a more reliable predictor of RESPONSE than the other. To explore trends related to the out-of-phase sequences, data Set D, was constructed from the response data for the duration-only subcondition in which *ba* was the longer syllable (manipulation level -1 in Figure 3), the glottalization-only condition, and the out-of-phase condition. The Set D results are charted in Figure 6. Here we see that the proportion of *baga* responses is highest in the glottalization-only condition, represented by the light grey columns. *Baga* responses were greatly decreased in all conditions in which *ba* was long. Informally, it appears that making the modal syllable longer washed out the influence of glottalization observed when no duration disparity was present.

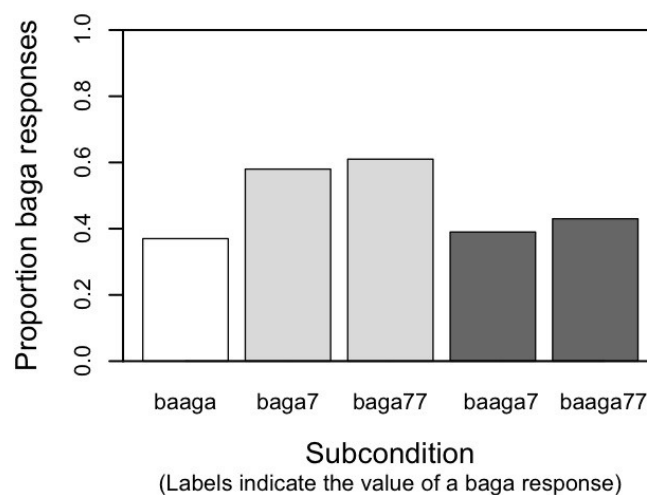


Figure 6: Proportion of *baga* responses in Set D. (White: duration-varying with long *ba*; Light grey: glottalization-only; Dark grey: Out-of-phase)

The Set D data was submitted to the same statistical analysis as described for Set C, with the exception that GA.DURATION was replaced with BA.DURATION. The best-fitting logistic model for the data as a whole was the model that contained only the intercept term (significant; coefficient = -0.4645, Odds = 0.63, SE = 0.163,  $z = -2.845$ ,  $p(>|z|) = 0.0044$ ). No other variable (whether predictor or design) contributed significantly to the model. The fact that the fixed effects of CREAK and BA.DURATION were not significant is consistent with our interpretation that they “cancelled one another out” in the out-of-phase condition.

To study the effect of glottalization in the conditions where *ba* was long, a data set (Set E) was constructed by removing the response data for the glottalization-only condition (light grey) from Set D, leaving only the response data for the conditions represented by the white and dark grey columns in Figure 6. The best-fitting model for the Set E data set was again the one that included only the intercept term (significant; coefficient = -0.640, Odds = 0.53, SE = 0.298,  $z = -2.148$ ,  $p(>|z|) = 0.0317$ ). Pairwise tests (Pearson’s Chi-square) indicated that there was no significant difference between the *baaga*, *бага7*, and *бага77* conditions being compared. In

other words, when *ba* was long, increasing glottalization on *ga* (from *ga* to *gaʔ* or *gaʔʔ*) did not significantly increase *baga* groupings, even though a weak trend seems to be evident in Figure 6.

To study effect of lengthening *ba* in the conditions where *ga* was creaky, the data for the glottalization-only (light grey) and out-of-phase (dark grey) conditions were combined and submitted to the same statistical test as was used for Sets A-E, with BA.DURATION as the predictor variable. Table 5, which contains the output of the best-fitting model, reveals that the effect of duration was highly significant: adding length to modal *ba* when *ga* was creaky lowered the odds of a modal-creaky grouping by 56%.

	Coef ( $\beta$ )	Odds	SE	z	P(> z )
INTERCEPT	0.424	1.53	0.143	2.965	= 0.003**
BA.DURATION	-0.818	.44	0.124	-6.951	< 0.00001***

Table 5: Output of the best-fitting statistical model for the response data in Set C.

Finally, the outcomes for sequences in the same glottalization classes are shown in Figure 7. Pairwise tests (Pearson's chi-square) indicated that the difference between the glottalization-only and out-of-phase conditions was significant at both manipulation levels (*bagaʔ* vs. *baagaʔ*:  $X^2 = 19.812$ ,  $df = 1$ ,  $p < 0.00001$ ; *bagaʔʔ* vs. *baagaʔʔ*:  $X^2 = 19.257$ ,  $df = 1$ ,  $p < 0.00001$ ). However, there were no significant differences between the glottalization-only and in-phase conditions, although the difference between *bagaʔ* vs. *bagaaʔ* approached significance (*bagaʔ* vs. *bagaaʔ*:  $X^2 = 3.287$ ,  $df = 1$ ,  $p = 0.06981$ ; *bagaʔʔ* vs. *bagaaʔʔ*:  $X^2 = 2.659$ ,  $df = 1$ ,  $p = 0.103$ ).

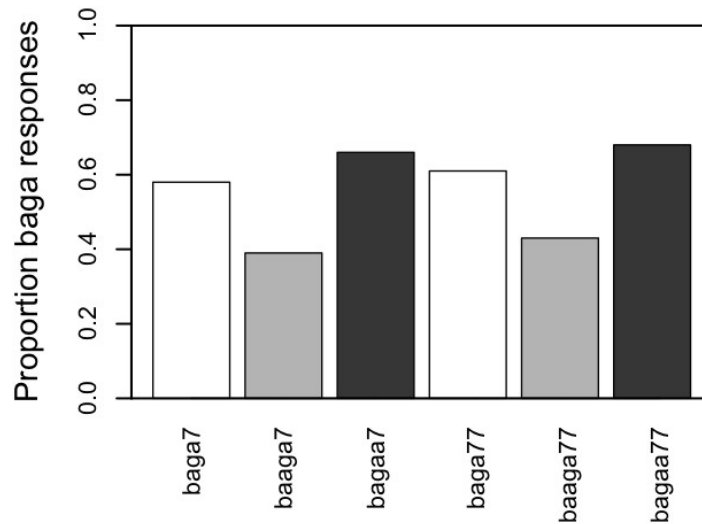


Figure 7. Proportion of *baga* responses for glottalized sequences. (White = glottalization only; light grey = out-of-phase; dark grey = in-phase)

These findings are interpreted in the discussion, which follows in section 3.

### 3 Discussion

In this study, we studied the perceptual influence of phonetic cues to duration and glottalization on English-speaking listeners' judgments of natural syllable grouping. The technique we used was to introduce systematic disparities in the duration and/or creakiness of the vowels in streams of recurrently alternating syllables, *ba* and *ga*. Our results indicate that disparities in duration produce the perception of short-long groupings, and disparities in glottalization produce the perception of modal-creaky groupings, as compared with a baseline condition in which neither feature was varied. These effects were highly significant in conditions in which these cues were varied singly. We further expected that when the duration of the glottalized period was increased, participants would be even more likely to choose groupings where the creaky syllable was in final position. While the data indicated a small trend in this direction, it did not reach significance in any condition in which glottalization was manipulated.

Our design included two conditions in which disparities in glottalization and duration were introduced into the same sequences. Sequences for an in-phase condition were prepared so that increased duration and glottalization marked the same syllable. In the case of the In-phase sequences, a recurrent short-long parse was also a modal-creaky parse, and the statistical analysis revealed a significant preference for this grouping. In this condition, the contributions of duration and glottalization seemed to be additive. Interpreting that statistical comparison of the outcomes in the In-phase, glottalization-only, the duration-varying condition in which *ga* was long but not glottal (see Figure 5), we found that adding glottalization to a long modal syllable significantly increased short-long groupings and that adding length to a creaky syllable also increased modal-creaky groupings. While both effects were statistically significant, the effect of glottalization was stronger.

The finding that glottalization may have contributed more to listeners' grouping decisions in the In-phase condition is in light of our findings for the second co-varied condition, in which duration and glottalization were manipulated "out of phase". In this condition, a short creaky syllable alternated with a long, modal one. The trends observed in the other conditions can be interpreted as indicating that the listeners preferred groupings in which the syllable with the demarcative feature came last. However, in the out-of-phase condition, these two grouping strategies were in competition: a short-long parse of an out-of-phase sequence was simultaneous a glottal-modal parse (and conversely, a modal-glottal parse was a long-short parse). A comparison between all conditions in which glottalization was manipulated indicated that adding length to the modal syllable significantly reduced modal-creaky grouping decisions, or increased short-long groupings. Interpreting the findings associated with co-varying glottalization and duration, it seems that when the two cues were working together in-phase, their contributions were additive, with glottalization contributing more. However, when the two cues were working against one another (out-of-phase), duration seems to have had a significant cancelling-out effect. Adding glottalization did not significantly increase modal-creaky grouping decisions when the modal syllable was long.

Summing up, our findings suggest at least two general outcomes that should be of interest for students of rhythm. The first is that vowel glottalization, a feature that has not been previously studied in a rhythmic context, tends to be heard as group-final. Glottalization has a demarcative function similar to that observed for duration, and the English-speaking listeners in our study used it similarly in judging natural syllable groupings. When both features are varied singly, the listeners in our study preferred groupings in which the syllable with the demarcative feature came last. The second finding of interest was that listeners seemed to use duration cues differently, depending on how duration was varied with another feature, in this case, glottalization. Interpreting our findings, in both of our co-varied conditions, increasing a duration disparity strengthened listeners' perception of short-long syllable groupings. However, the effect of duration was much stronger when it worked against glottalization than when the two features worked together. These discoveries indicate that much can be learned from extending the study of rhythmic grouping preferences to include acoustic features other than intensity and duration, and from exploring the relative contributions of these features by testing more complex feature combinations.

#### 4 References

- Beckman, M. E., and J. Edwards. 1994. Articulatory evidence for differentiating stress categories. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, ed. P. A. Keating, 7-33. Cambridge University Press.
- Bhatara, Anjali, Natalie Boll-Avetisyan, Annika Unger, Thierry Nazzi, and Barbara Höhle. 2013. Native language and stimulus complexity affect rhythmic grouping of speech. *Journal of the Acoustical Society of America*.
- Bion, R. A. H., S. Benavides-Varela, and M. Nespor. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech* 54:123-140.
- Boersma, Paul, and David Weenink. 2011. Praat: doing phonetics by computer [Computer program]. Version 5.3.03. <http://www.praat.org/>.
- Bolton, Thaddeus L. 1894. Rhythm. *American Journal of Psychology* 6 (2):145-238.
- Byrd, D., J. Krivokapi, and S. Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120:1589-1599.
- Campbell, N., and M. Beckman. 1997. Accent, stress and spectral tilt. *Journal of the Acoustical Society of America* 101(5):3195.
- Crowhurst, Megan. 2013. Perceptual influences on rhythmic grouping preferences. In *Talk presented at M@90: A workshop on metrical structure in honor of Morris Halle's 90<sup>th</sup> birthday*. Cambridge, MA.
- Crowhurst, Megan, and Amador Teodocio. In press. The joint influence of duration and intensity on the subjective grouping of rhythmic speech among Zapotec and English speakers. *Phonology*.

- Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24:423-444.
- Fletcher, J. 2010. The Prosody of Speech: Timing and Rhythm. In Hardcastle W., Laver J. & Gibbon F. (eds.). *The Handbook of Phonetic Sciences*. United Kingdom: Wiley-Blackwell Publishing, pp. 523-602.
- Frazier, Melissa. 2009. The Production and Perception of Pitch and Glottalization in Yucatec Maya. Doctoral Dissertation, University of North Carolina at Chapel Hill.
- Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress, *Journal of the Acoustical Society of America* 27, 765–768.
- Fry, D. B. 1958. Experiments in the perception of stress, *Language and Speech* 1, 126–152.
- Gay, T. 1978. Physiological and acoustic correlates of perceived stress. *Language and Speech* 21(4):347-353.
- Gerfen, C., and K. Baker. 2005. Production and perception of laryngealized vowels in Coatzacoapan Mixtec. *Journal of Phonetics* 33(3):311-334.
- Gordon, M., and P. Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29:383-406.
- Gussenhoven, C., and A. C. M. Rietveld. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20:283-303.
- Halle, M. 1973. Stress rules in English: a new version. *Linguistic inquiry*, 451-464.
- Hay, Jessica, and Randy Diehl. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics* 69:113-122.
- Hayes, B. 1980. *A Metrical Theory of Stress Rules*. PhD dissertation, MIT, Cambridge Massachusetts.
- Hayes, Bruce. 1995. *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
- Huber, D. 1988. Aspects of the communicative function of voice in text intonation. Doctoral Dissertation, Chalmers University.
- Iversen, John R., Aniruddh D. Patel, and Kengo Ohgushi. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America* 124:2263-2271.
- Kusumoto, K., and E. Moreton. 1997. Native language determines parsing of nonlinguistic rhythmic stimuli. *Journal of the Acoustical Society of America* 102 (5):3204.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: The MIT Press.
- Pierrehumbert, J. 1995. Prosodic effects on glottal allophones. In *Vocal fold physiology: voice quality control*, 39-60. San Diego: Singular Publishing Group.
- Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14, 19-100.
- Redi, Laura, and Stefanie Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 407-429:29.
- Rice, Curtis. 1992. *Binarity and ternarity in metrical theory: Parametric extensions*. PhD dissertation, the University of Texas at Austin.



- Streeter, Lynn A. 1978. Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America* 64:1582-1592.
- Trainor, Laurel J., and Beth Adams. 2000. Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception and Psychophysics* 62:333-340.
- Urbanek, Simon, Hans-Jörg Bibiko, and Stefano M. Iacus. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, ISBN 3-900051-07-0.
- Woodrow, Herbert. 1909. A Quantitative Study of Rhythm: The effect of variations in intensity, rate, and duration. *Columbia University Contributions to Philosophy and Psychology*, Vol. XVIII, No. 1.

## Appendix A: Example of a response sheet

### **Block 1**

1.	... ba (ga ba) ga ba ...	How sure?	yes	no
2.	... ga ba ga (ba ga) ba ...	How sure?	yes	no
3.	... ga ba ga ba ga ...	How sure?	yes	no
4.	... ba ga ba ga ...	How sure?	yes	no
5.	... ba ga ba ga ba ga ...	How sure?	yes	no
6.	... ga ba ga ba ...	How sure?	yes	no
7.	... ba ga ba ga ba ...	How sure?	yes	no
8.	... ga ba ga ba ga ba ...	How sure?	yes	no
9.	... ga ba ga ba ga ...	How sure?	yes	no
10.	... ba ga ba ga ba ...	How sure?	yes	no
11.	... ba ga ba ga ba ga ...	How sure?	yes	no
12.	... ga ba ga ba ...	How sure?	yes	no

## Appendix B: Response data

	-1	0	1
<i>bag</i>	159 (.56)	117 (.41)	106 (.37)
<i>a</i>			
<i>gab</i>	127 (.44)	171 (.59)	181 (.63)
<i>a</i>			

Table 6: Response data for control and duration-only sequences (Set A). The scale reflects the duration of *ba* relative to *ga*.

	0	1	2
<i>bag</i>	117 (.41)	166 (.58)	176 (.61)
<i>a</i>			
<i>gab</i>	171 (.59)	119 (.42)	111 (.39)
<i>a</i>			

Table 7: Response data for control and glottalization-only sequences (Set B). The scale reflects increases in the glottal quality of *ga* (0 = *ga*, 1 = *gaʔ*, 2 = *gaʔʔ*).

	<i>In-phase</i>		<i>Out-of phase</i>	
	<i>bagaaʔ</i>	<i>bagaaʔʔ</i>	<i>baagaʔ</i>	<i>baagaʔʔ</i>
<i>bag</i>	188 (.66)	195 (.68)	111 (.39)	122 (.43)
<i>a</i>				
<i>gab</i>	97 (.34)	91 (.32)	172 (.61)	164 (.57)
<i>a</i>				

Table 8: Response data for in-phase sequences (included in Set C) and out-of-phase sequences (included in Set D). Column headings indicate properties of groupings represented by a *baga* response.